APPENDIX A

Attorney Docket 11910-002001

Appendix A

Three aspects of the invention are:

1. The recognition of the desirability of displaying to a financial investment customer in real time, for example on a World Wide Web site, the probability distribution governing the price of a particular asset (*e.g.*, a stock) at a selected future time.

2. The recognition that such probability distributions can be derived from option prices for that asset, or for related assets, which are readily available in real time.

3. The recognition that probability distributions involving several asset prices simultaneously are useful to investment customers in several contexts, especially in exploring hypothetical scenarios, and that single asset distributions such as (but not restricted to) the above can be meaningfully incorporated into multivariate distributions, manageably determined.

In this appendix we first describe a basic method for deriving probability distributions for single assets from option prices. We next describe improvements on this basic method to address various practical issues. Then we take up the multivariate case and show how to extend this kind of single asset price distribution, or any other, to the multivariate case. Finally, we consider a number of novel multivariate applications, with emphasis on scenario exploration.

1 Basic method

A call option is an option to buy an asset (*e.g.*, a stock) at a certain price x (called the strike price) on a given expiration date T days in the future. (An option exercisable only on the expiration date is called a European-style option; for simplicity we will consider in this discussion only this type of option.[1]) Similarly, a put option is an option to sell an asset at a strike price x on a given expiration date. (The "European-style" assumption of no possible early exercise is more important here, but can also be ignored for puts that are not too deeply "in the money.")

Let $c(x)$ denote the price of a call option on an asset at strike price x, and $p(x)$ the price of a put option. Such prices are established by options market-makers. We have realized that such prices implicitly contain information about a "market view" of the probability distribution of the price of that asset at the expiration date.

In a simple but precise form, this market view can be stated as follows. Suppose that we were given the call price curve $c(x)$ or the put price curve $p(x)$ as a continuous function of the strike price x for all $x > 0$. Then, the second derivative of either the call or the put price curve is the market view of the risk-neutral probability density function (pdf) $f(x)$ of the asset price at the expiration date. In other words, $f(x) = c''(x) = p''(x)$.

The idea that option prices determine some kind of implied probability distribution is fairly well known in the financial literature. The idea that a pdf can be computed by taking the second derivative of a continuous option price curve is known in the academic literature, but it does not appear to be very well known. For example, the standard textbook "Options, Futures,

[1] Even allowing for possible early exercise, most liquidly traded call options without large dividends can be treated as if there were no possibility of such exercise, since sale of the option is usually a better alternative; therefore, these call options behave similarly to European-style options.

and Other Derivatives," by John C. Hull (Fourth Edition, 1999; Prentice-Hall) mentions implied probabilities, but not the second-derivative method. The best reference that we have been able to find is J. C. Jackwerth and M. Rubinstein,"Recovering probability distributions from option prices," *J. Finance*, vol. 51, pp. 1611–1631 (1996), which has only six prior references.

The risk-neutral distribution (at a fixed future time T, for a fixed asset) is defined as the price distribution that would hold if market participants were neutral to risk, which they generally are not. However, many asset pricing theories, such as those underlying Black-Scholes option theory and most of the variations found in the Hull book above, allow for the true risk-averse asset price distribution to be obtained from the risk-neutral distribution $f(x)$ just by adjusting the latter by an appropriate risk premium: If there are no dividends, the true distribution is just $f(xe^{(\mu-r)T})$, where $\mu - r$ is the expected annual return rate for the stock in excess of the risk free rate r. We use a variation on this simple format, slightly modified to allow for dividends (see below), though our invention could also work well with a more complicated adjustment. In this format, a value for $\mu - r$ must still be supplied. We use as a default the "consensus estimate" taken from the textbook "Active Portfolio Management" (1995) by Grinold and Kahn. These authors note a long-term average value of the risk premium to be 6% per year, and suggest multiplying this number by the stock's beta to get $\mu - r$. The parameter beta is the slope of the line giving a regression of the stock in question against a market portfolio, often taken as the S&P 500. This is the well-known CAPM estimate for the expected excess return. Whether good or bad, its stature as a consensus estimate makes it suited to our aim of providing a market view, though it is only a default. Our invention, which provides the risk-neutral component of the probabilities, could work with other estimates for the risk-averse adjustment parameter $\mu - r$ and with any explicit scheme for adjusting the risk neutral probability density to the risk-averse probability density. It is worth pointing out that, for shorter time periods-even a month or two–the risk adjustment required is small and generally overwhelmed by fluctuations in the risk-neutral distribution itself.

We give a brief proof that the second derivative procedure gives the correct risk-neutral probability distribution. As in Hull, we may calculate the European call or put price as an expected value in the risk-neutral distribution.

If the actual value of the asset on the expiration date is v, then the value of a call option at strike price x is $\max\{v - x, 0\}$, and the value of a put option is $\max\{x - v, 0\}$. If the actual value is a random variable with pdf $f(v)$, then the expected value of a call option at x at the expiration date is

$$c_T(x) = \mathsf{E}_v[\max\{v - x, 0\}] = \int_x^\infty (v - x) f(v)\, dv,$$

and the expected value of a put option at x at the expiration date is

$$p_T(x) = \mathsf{E}_v[\max\{x - v, 0\}] = \int_0^x (x - v) f(v)\, dv.$$

The current values $c(x)$ and $p(x)$ may be obtained by discounting $c_T(x)$ and $p_T(x)$ by e^{-rT}, where r is the risk-free interest rate, but for our purposes, forecasting probability distributions at time T, we do no discounting, and henceforth just write $c(x) = c_T(x)$, $p(x) = p_T(x)$.

Parenthetically, from these expressions we observe that

$$p(x) - c(x) = \int_0^\infty (x - v) f(v)\, dv = x - E_v[v] = x - s^*,$$

where $s^* = E_v[v]$ is the expected value of the asset at the expiration date under the risk-neutral distribution. (If there are no dividends, then $s^* = se^{rT}$; if there are dividends, then in general it is necessary to subtract from se^{rT} the value at time T of the dividends.) This well-known relation is called put-call parity; it shows why either price curve carries the same information.

From the above expression for $c(x)$, it follows that its first derivative is

$$c'(x) = -\int_x^\infty f(x)\, dx = F(x) - 1,$$

where $F(x) = \int_0^x f(v)\, dx$ is the cumulative distribution function (cdf) of the random variable v. To prove this, note that $v - x = \int_x^v dz$. Therefore

$$c(x) = \int_x^\infty (v-x) f(v)\, dv = \int_x^\infty dv \int_x^v dz\ f(v) = \int_x^\infty dz \int_z^\infty dv\ f(v) = \int_x^\infty dz\ (1 - F(z)),$$

where we interchange the variables v, z to integrate over the two-dimensional region $\mathcal{R} = \{(v, z) : x \le z \le v\}$. The last expression implies that $c'(x) = -(1 - F(x))$.

From put-call parity, it follows similarly that

$$p'(x) = 1 + c'(x) = F(x).$$

Since the cdf and pdf are related by $F'(x) = f(x)$, these expressions in turn imply that the second derivative of either $c(x)$ or $p(x)$ is the pdf $f(x)$:

$$c''(x) = p''(x) = F'(x) = f(x).$$

The general character of the option price curves $c(x)$ and $p(x)$ is therefore as follows:

- For all x less than the minimum possible value of v (*i.e.*, such that $F(x) = 0$), $c(x) = \mathsf{E}_v[v] - x = s^* - x$ and $p(x) = 0$. In other words, $c(x)$ is a straight line of slope -1 starting at $c(0) = \mathsf{E}_v[v] = s^*$, while $p(x) = 0$.
- For all x greater than the maximum possible value of v (*i.e.*, such that $F(x) = 1$), $c(x) = 0$ and $p(x) = x - s^*$. In other words, $p(x)$ is a straight line of slope $+1$ and x-intercept s^*, while $c(x) = 0$.
- These two line segments are joined by a continuous convex $\cup$ curve whose slope increases from -1 to 0 for $c(x)$, and from 0 to $+1$ for $p(x)$.

We note that the fact that the mean $\mathsf{E}_v[v]$ of the pdf $f(x)$ is s^*, the value in future dollars at time T of the underlying price s (less the value of any dividends), implies that option prices must be constantly adjusted to reflect changes in the underlying price s, even if there is no market activity in the options.

The fact that $s^* = \mathsf{E}_v[v]$ also implies that an option price curve can make no prediction about the general direction of the underlying price s. However, the option price curve does predict the shape of the pdf $f(x)$, and in particular its volatility.

1.1 Approximations based on a finite subset of bid-asked option prices

In practice, option prices $c(x)$ and $p(x)$ are quoted only for a finite subset of equally-spaced strike prices x, namely $x = n\Delta$ for integer n and spacing Δ. We denote $c(n\Delta)$ and $p(n\Delta)$ by c_n and p_n, respectively. Moreover, quotes specify only a bid-asked spread, not exact prices. In this subsection we give methods for dealing with these problems. (Most of the Jackwerth–Rubinstein paper (*op. cit.*) is concerned with these kinds of curve-fitting problems.)

The first derivatives $c'(x)$ and $p'(x)$ at $x = (n + \frac{1}{2})\Delta$ may be estimated by the first differences

$$\hat{c'}_{n+\frac{1}{2}} = \frac{c_{n+1} - c_n}{\Delta}; \quad \hat{p'}_{n+\frac{1}{2}} = \frac{p_{n+1} - p_n}{\Delta}.$$

The corresponding estimates of the cdf $\hat{F}_{n+\frac{1}{2}} = F((n + \frac{1}{2})\Delta)$ are

$$\hat{F}_{n+\frac{1}{2}} = 1 + \hat{c'}_{n+\frac{1}{2}}; \quad \hat{F}_{n+\frac{1}{2}} = \hat{p'}_{n+\frac{1}{2}}.$$

Thus, using both bid and ask prices for both European-style puts and calls, one can compute four different estimates for the cdf $\hat{F}_{n+\frac{1}{2}}$, which can then be combined into a single estimate. This combination will preferably take into account whether $x = (n + \frac{1}{2})\Delta$ is much less than the underlying price s ("deep out-of-the-money"), near s ("near the money"), or much greater than s ("deep in-the-money"), according to the different patterns of setting bid-asked spreads in these different ranges. Another consideration is avoiding quotes near prices where early exercise is likely, such as deep in-the-money puts.

Similarly, the second derivatives $c''(x)$ and $p''(x)$ at $x = n\Delta$ may be estimated by the first differences of the estimates of the first derivatives; *e.g.*,

$$\hat{c''}_n = \frac{\hat{c'}_{n+\frac{1}{2}} - \hat{c'}_{n-\frac{1}{2}}}{\Delta} = \frac{c_{n+1} - 2c_n + c_{n-1}}{\Delta^2}.$$

We may take $\hat{c''}_n$ or $\hat{p''}_n$ or some combination as above as our estimate $\hat{f}_n$ of the pdf $f(n\Delta)$.

Note that since $f(x) \geq 0$, option prices should satisfy a convexity condition, *e.g.*, $c_{n+1} - 2c_n + c_{n-1} \geq 0$ for call option prices. Indeed, violation of this condition would allow making money via a risk-free "butterfly straddle" involving buying one call option at $(n+1)\Delta$ and another at $(n-1)\Delta$, and selling two call options at $n\Delta$. A similar result holds for put options.

1.2 Dynamic estimates

The methods considered in the previous subsection allow estimation of the cdf and pdf at a subset of Δ-spaced values of x, based on a static set of option quotes at a particular time.

As previously noted, however, option prices must change continually in response to changes in the underlying price s. Let s^* denote the corresponding forward price at expiration (the price s evaluated with interest). Suppose this price (measured in dollars at expiration) moves up (or down) by a small amount, an increment ϵ in its logarithm, say, with little or no change in volatility. Here ϵ may be viewed as, approximately, the percentage move δ/s^* caused by a move of δ in the (forward) stock price. We expect in this situation that (forward) probability distribution for the stock price will just be shifted by ϵ in the log domain. That is, the distribution will appear to be identical there, except with a mean shifted by ϵ. Thus, the value of the new cdf at $x = e^{\ln x}$ is $F(e^{(\ln x - \epsilon)}) = F(x/a)$, where F denotes the original cdf with distribution mean

s^*, and $a = e^\epsilon$. A reasonable call price functional equation that gives the same effect, upon differentiation, is

$$ac(s^*, x/a) = c(as^*, x),$$

where $c(s^*, x)$ denotes the price, in dollars at expiration, for a call option at strike x when the underlying is at price s. Note in this equation that all other variables, such as volatility, are assumed to be the same, which will only be approximately true, even for very small values of ϵ.

But, assuming this approximation, we can think of an option price at strike x, measured when the (forward) price has moved to as^*, for a near 1, as giving instead a times the price of an option at strike x/a, but corresponding to the current underlying price s^*. Considering all the strikes at which options are frequently quoted, and thinking additively, we can effectively observe $c(x)$ (and $p(x)$) for a different subset of approximately equally-spaced strike prices, roughly $x = n\Delta - \delta$ for various values of $\delta = \epsilon s^*$. Some care must, of course, be taken to ensure simultaneity of prices, of option and underlying. For this reason, we may prefer to consider the values of $n\Delta$ (corresponding to the various standard strike values) separately, and synchronize observed time of sales for an option at a given strike with the underlyling security. Implied volatilities (discussed below) could be monitored, to ensure their changes relative to ϵ were small.

Using a similar technique to that described in the above paragraphs, meaningful average option prices for a given strike can also be computed, using thin strike intervals and using either short time intervals or time series methods (time averages weighting the present more than the past). Note that, without the framework described in this subsection, the computation of "average" option prices at a given strike are problematic when the stock price varies in the period over which the average is taken.

To summarize: Given enough movements of the underlying price, we can effectively observe prices and compute estimates as above for a much more finely quantized subset of strike prices x, and provide a framework for improving accuracy through averaging methods.

2 Methods for extrapolation and smoothing

There are two major limitations to the basic methods of the previous section. One is that option quotes are available only for certain expiration dates. Another, not so obvious, is that option quotes are reliable primarily for options in which there is substantial market activity. These would typically be nearer-term options at strike prices near the money (the underlying price).

To extend our prediction methods to times other than expiration dates and over wider ranges of strike prices (and also to help reduce "noise" in our displays), we use extrapolation and smoothing techniques. We have found that it is advantageous to do extrapolation and smoothing in the volatility domain.

There are many reasons for this advantage. For example, option practitioners are well aware of the kinds of shapes that the volatility curves (sometimes called "volatility smiles") have had historically, in various markets, and how these curves vary with time; this can be a guide to imposing structure on the smoothing curves to prevent overfitting of possible artifacts. Many records have been kept of the volatilities implied by option prices, and it is easy to examine how in the past they have changed with respect to price behavior. For example, the Chicago Board of Options Exchange makes public its average near-the-money volatility index (now called VIX) for S&P 100 options back to 1986. Finally, it is easier to work visually with volatility curves, which

would theoretically be flat if $f(x)$ were lognormal, than with visual differences in near-lognormal pdfs, which can all look very much alike. Mathematically, model improvements can be made in the volatility domain just by changing coefficients of low-degree polynomial approximations, even though these affect higher-order terms in power series for the corresponding cdfs or pdfs.

The following subsections explain more precisely how to work in the volatility domain.

2.1 Lognormal pdfs

The standard Black-Scholes theory of option pricing (see Hull, *op. cit.*) yields a lognormal pdf $f(v)$ whose expected value is $\mathsf{E}_v[v] = s^*$, such that $\ln v$ is a Gaussian (normal) random variable with variance $\sigma^2 T$, where the parameter σ is called the volatility rate of the asset, and T is the time to expiration. By a standard property of lognormal distributions, this implies that the mean of $\ln v$ is $\mathsf{E}_v[\ln v] = \ln s^* - \sigma^2 T/2$.

From this pdf follows the famous Black-Scholes call option pricing formula [Hull, Appx. 11A]:

$$c(x) = \mathsf{E}_v[\max\{v - x, 0\}] = s^* N(d_1(x)) - xN(d_2(x)),$$

where $N(d_1(x))$ and $N(d_2(x))$ are values of the cumulative distribution function of a Gaussian random variable of mean zero and variance 1 at the points

$$\begin{aligned} d_1(x) &= \frac{\ln(s^*/x) + \sigma^2 T/2}{\sigma\sqrt{T}} = d_2(x) + \sigma\sqrt{T}; \\ d_2(x) &= \frac{\ln(s^*/x) - \sigma^2 T/2}{\sigma\sqrt{T}} = \frac{\mathsf{E}_v[\ln v] - \ln x}{\sigma\sqrt{T}}. \end{aligned}$$

(Recall that our version of the call price is not discounted, and is given in dollars at time T, and that s^* is today's stock price, valued in dollars at time T, less the value of any dividends.) Note that $\sigma\sqrt{T}$ is the standard deviation of $\ln v$; therefore $-d_2(x)$ is just $\ln x$, measured in standard deviations from the mean $\mathsf{E}_v[\ln v]$.

Similarly, by put-call parity, we have the Black-Scholes put option pricing formula

$$p(x) = c(x) + x - s^* = s^*(N(d_1(x)) - 1) - x(N(d_2(x)) - 1) = xN(-d_2(x)) - s^* N(-d_1(x)).$$

Taking the derivative with respect to x, and using $s^* N'(d_1(x)) = xN'(d_2(x))$ and $d_1'(x) = d_2'(x)$ (the latter equation holding under the assumption of constant volatility, which we will later drop), we obtain

$$F(x) = c'(x) + 1 = -N(d_2(x)) + 1 = N(-d_2(x)).$$

Now $F(x)$ is the probability that $v \le x$, which is equal to the probability that $\ln v \le \ln x$, which since $\ln v$ is Gaussian with mean $\mathsf{E}_v[\ln v]$ and standard deviation $\sigma\sqrt{T}$ is given by

$$F(x) = \Pr\{v \le x\} = \Pr\{\ln v \le \ln x\} = N\left(\frac{\ln x - \mathsf{E}_v[\ln v]}{\sigma\sqrt{T}}\right) = N(-d_2(x)).$$

Thus we have verified that the Black-Scholes pricing formulas give the correct cdf $F(x)$. The derivative of $F(x)$ will thus yield the correct lognormal distribution $f(x) = F'(x)$.

2.2 Characterization of general cdfs

Now let $F(x)$ be an arbitrary cdf on $\mathbb{R}_+$; *i.e.*, a function that monotonically increases from 0 to 1 as x goes from 0 to infinity. For simplicity we will assume that $F(x)$ is strictly monotonically increasing; *i.e.*, $f(x) = F'(x) > 0$ everywhere. Then there exists a continuous one-to-one "warping function" $y : \mathbb{R}_+ \to \mathbb{R}$ such that $F(x) = N(y(x))$ everywhere; *i.e.*, such that the probability that a random variable v with cdf $F(x)$ will satisfy $v \leq x$ is equal to the probability that a standard Gaussian random variable n with mean zero and variance 1 will satisfy $n \leq y(x)$. Similarly, there is an inverse warping function $x(y)$ such that $F(x(y)) = N(y)$.

Given the warping function $y(x)$, the cdf $F(x)$ may be retrieved from the relation $F(x) = N(y(x))$. Therefore the cdf $F(x)$ completely specifies the warping function $y(x)$, and *vice versa*; *i.e.*, both curves carry the same information.

If $F(x)$ is the cdf of a lognormal variable v such that $\ln v$ has mean $\mathsf{E}_v[\ln v] = \ln s^* - \sigma^2 T/2$ and variance $\sigma_v^2 = \sigma^2 T$, as in the previous subsection, then the warping function is given by

$$y(x) = -d_2(x) = \frac{\ln x - (\ln s^* - \sigma^2 T/2)}{\sigma\sqrt{T}} = \frac{\ln x - \mathsf{E}_v[\ln v]}{\sigma_v}.$$

For this reason we may sometimes write $y(x)$ as $-d_2(x)$, even when the cdf is not lognormal so that the right-hand equation above for $d_2(x)$ does not hold.

2.3 Implied volatilities

If $f(x)$ is not lognormal, then the Black-Scholes pricing formulas do not hold. Nonetheless, given an option price $c(x)$ or $p(x)$, it is common practice to define the implied volatility $\sigma(x)$ as the value of σ such that the Black-Scholes pricing formula holds, for a given x, s and T.

The implied volatility curve $\sigma(x)$ so defined is a function of the strike price x, which is constant if and only if the pdf $f(x)$ is actually lognormal. In practice, it is typically a convex $\cup$ curve, called a "volatility smile." See, *e.g.*, Hull, Chapter 17.

From Subsection 2.1, we can see that there is a second method of calculating implied volatilities, as follows. Suppose that we have an estimate of the cdf $F(x)$. Define the cdf-implied volatility $\sigma_1(x)$ as the value of σ such that the Black-Scholes cdf formula $F(x) = N(-d_2(x, \sigma, T))$ holds, for a given x, s and T.

The first method has the advantages of being defined directly from raw price data, and of being well understood in the financial community. However, the second method has the following advantages:

1. It is easier to calculate, at least from estimates of $F(x)$;

2. It gives a simpler and arguably more intuitive relationship between volatility and the cdf $F(x)$. If we use the traditional implied volatility $\sigma(x)$, then the relationship is instead

$$F(x) = N(-d_2(x)) + \frac{\partial c}{\partial \sigma}\sigma'(x).$$

3. It fits better with the multivariate theory to be developed below.

We have observed that the two curves $\sigma(x)$ and $\sigma_1(x)$ seem to be fairly similar, at least as to the direction of their slope, and are generally not too far apart in value "near the money". Also $\sigma_1(x) = \sigma(x)$ whenever $\sigma(x)$ has zero slope, though $\sigma_1(x)$ is a little smaller than $\sigma(x)$ when the slope $\sigma'(x)$ is negative (which often occurs for stocks). See the above equation. Finally, one function is as *ad hoc* as the other. Therefore, because of the above reasons, we generally prefer to use the cdf-implied volatility curve $\sigma_1(x)$.

In any case, it is clear that either $\sigma(x)$ or $\sigma_1(x)$ contains the same information as any of the curves $c(x)$, $p(x)$, $F(x)$ or $f(x)$. From $\sigma(x)$ or $\sigma_1(x)$ we can recover $c(x)$ or $F(x)$ using the Black-Scholes call option pricing or cdf formula, and from this we can obtain all other curves.

2.4 Extrapolation and smoothing in the volatility domain

The volatility curve $\sigma(x)$ or $\sigma_1(x)$ may be calculated pointwise from the corresponding curve $c(x)$ or $p(x)$ to give a set of values at a finite subset of strike prices x. Each of these values may be deemed to have a certain degree of reliability.

It is then a standard problem to fit a smoothed and extrapolated curve $\tilde{\sigma}(x)$ or $\tilde{\sigma}_1(x)$ to these points, taking into account their relative reliabilities. Any standard smoothing and extrapolation method may be used. In general, the usual problems of avoiding overfitting or oversmoothing must be addressed.

It is well-known that implied volatilities also vary with time. We generally wish to estimate curves $\tilde{\sigma}(x,T)$ or $\tilde{\sigma}_1(x,T)$ as replacements for the constant volatility σ in the Black-Scholes formulas, *e.g.*, $c(x) = c(x,\sigma,T)$ or $F(x) = N(-d_2(x,\sigma,T))$.

In an especially meaningful example, we have experimented with a class of smoothing algorithms used in "Implied volatility functions: Empirical tests," by B. Dumas, J. Fleming and R. E. Whaley, *J.Finance*, vol. 53, pp. 2059-2106, Dec. 1998. These authors fit an implied volatility curve $\sigma(x)$, for the purpose of setting up a "strawman" option price model for testing (and defeating) a theory regarding the role of volatility in option pricing. Their "strawman" option pricing model $c(x)$ was obtained by putting the resulting smoothed curve back into the Black-Scholes call formula. It is a "strawman" *ad hoc* model, because no intuitive notion of stock volatility could possible vary with strike price, which the stock never "sees." Nevertheless, their model performed admirably, surpassing in predictive power the highly regarded "implied tree" method. One possible explanation offered was that their model mimicked in a smooth way interpolation methods actually employed by practitioners in the options markets. (See the discussion of "Volatility matrices" in Hull, cited above.) Such an approach to option pricing seems ideal to us, because of its accuracy and because its underlying rationale represents a market view. Thus, we use the Dumas-Fleming-Whaley model for our own entirely different purpose, that of forecasting probability distributions. All that is necessary is to differentiate their call price model, which, conveniently for us, is a smooth function of strike price and other standard variables such as time, current stock price, and the risk-free rate of interest. The formula for the cdf $F(x)$ is, as before, this derivative with 1 subtracted, or

$$F(x) = N(-d_2(x)) + \frac{\partial c}{\partial \sigma}\sigma'(x).$$

We can make this very explicit. We have

$$\frac{\partial c}{\partial \sigma} = x\sqrt{T}N'(-d_2(x))$$

where $N'(z)$ denotes standard normal density, while $\sigma'(x)$ may be computed by differentiating the Dumas-Fleming-Whaley fitted volatility curve. The latter has the form

$$\sigma(x,T) = a_0 + a_1 x + a_2 x^2 + a_3 T + a_4 T^2 + a_5 xT.$$

The coefficients $\{a_i\}$ are determined by regression. This kind of quadratic curve-fitting is easily implemented. Dumas-Fleming-Whaley impose a constraint to prevent their volatilities from going below 0 (or even below 0.01), and we have imposed further constraints on extrapolations (which we often carry out beyond the range of their tests), to ensure that the final cdf does not go below zero or above one. We have experimented with other variations on their basic approach, for example, using linear interpolation in the time domain, where we do not need to take deivatives. Our methods would, of course work, with any approach, possibly quite different, to volatility curve-fitting, though the general Dumas-Fleming-Whaley approach has many things going for it: accuracy, conformity to marketplace use of Black-Scholes, smoothness (differentiability, in particular), conformity to historical experience regarding the smile structure of volatility curves (especially important for extrapolation), and simplicity (which, beyond ease of implementation, helps avoid overfitting). These advantages are achieved in a probability context that was not considered in the paper where these volatility curves were introduced.

3 The multivariate case

The methods in the previous sections are capable of generating a display of raw or smoothed and extrapolated probability distributions for any optionable asset. Option prices are quoted on a large number of securities, as well as on certain indices, such as the S&P 500.

However, an investor would also like to know future probability distributions for:

- His or her entire portfolio;
- Mutual funds;
- A security without a quoted option;
- A security in a hypothetical scenario.

All of these questions involve considerations of several securities at once, and the probabilities of their simultaneous configuration of prices. This is clearly a consideration in the first two items above, but also enters in the third, where we would want to extract as much information as possible about the security without a quoted option price from those correlated with it that do have quoted options. Finally, in scenario analysis there are many questions that involve considering the probabilities of several security prices occurring at once, including changes in factors influencing the market that might be modeled by changes in a portfolio of those securities most affected. We will take all of these issues up in the remainder of this document, but for now we just try to give a basic introduction.

For a portfolio of securities, or a mutual fund, we are interested in a composite asset of the form

$$x = h_1 x_1 + h_2 x_2 + \cdots + h_n x_n,$$

where the x_i are all assets for which we individually know the cdf $F(x_i)$ or the pdf $f(x_i)$. To give our method the most flexibility, we do not require that this knowledge come from any particular

procedure, though we favor the approach of the preceding two sections. However, even for some securities or indices with a quoted option, we might not feel there was sufficient option activity to justify a full fit of a volatility curve, and might take a cruder substitute, even a flat straight line based on an average of available implied volatilities. In addition, it is convenient to allow the possibility that a few assets we are monitoring might not have any quoted option at all; this is easily accommodated by, say, using a flat volatility curve with a historical value for volatility. For testing purposes and comparisons we might even want to consider a list of assets with all volatility curves given this way. In any case our methodology here is very general, and we only require that we know warping functions $y_i(x_i)$ such that $F(x_i) = N(y_i(x_i))$ for all i. If the asset has an active options market, then the warping functions may be determined by either first estimating $F(x_i)$ directly from (finite differences of) options price data, as in subsection 2.2, or by using the approach discussed later in Section 2 of extrapolating and smoothing in the volatility domain. In the latter case we have an explicit form of the warping function $y_i(x_i)$ in terms of a fitted volatility curve $\tilde{\sigma}_1(x_i, T)$ as $y_i(x_i) = -d_2(x_i, \tilde{\sigma}_1(x_i, T), T)$, and this equation can also be used with any volatility curve with the assets above, that might have fewer or no traded options. In a later section we will discuss portfolios in the logarithm domain, possibly containing long and short positions. One can think of warping them to standard normal directly, subtracting the mean and dividing by the standard deviation. Alternately, to keep our notation uniform, one can invent an asset with price x_i such that $-d_2(x_i)$ gives this warped value (using for $\sigma_1(x)$ the observed historical volatility). But we wish to emphasize that the method we are describing works with ANY single-variable warping functions, even using a different one for each variable. The only further substantitive ingredient is the plausibility of using JOINTLY normal distribuitions, which we now discuss.

The general problem is to find a multivariate probability distribution for the complete set of variables $(x_1, \ldots, x_n)$, or equivalently for their logarithms. In simple financial models generalizing the Black-Scholes framework, the multivariate distribution of the logvariables is multivariate (*i.e.*, jointly) normal; see Musiela and Rutkowski's book "Martingale methods in financial markets" (1999). This implies that all portfolios of these logvariables are jointly normal, and can also be used with other logvariables and portfolios of them to form a jointly normal distribution. Thus, if we wish, it is reasonable to use BARRA (or functionally equivalent) factors as single (log)variables in our model, using, say, individual normal distributions for them based on historical volatility. These factors may represent fundamentals of companies or even macroeconomic variables such as interest rates. We do not further discuss such factors, but refer to the book of Grinold and Kahn cited above, which also describes how to closely approximate them as portfolios of security returns. Our preference is to not use BARRA factors directly, but stay as much as possible in the world of optionable securities, and address questions involving BARRA factors in terms of approximating portfolios consisting mostly of optionable securities. (But for testing and comparisons, it is still useful to be able to include them directly, and we do have that capability.)

Now we certainly do not wish to use only the simple multidimensional Black-Scholes model, which would not directly allow the nonlognormal input from our single-variable distributions based on the options markets. At the same time, option prices on individual assets do not tell us anything about how assets interact, in particular, their correlations. Fortunately, correlations may be estimated from past (historical) data, and may be viewed as covariances for data that has been standardized (has standard deviation 1). Each multivariate normal distribution is determined by its mean and covariance matrix. Thus, a natural approach is to use the individual distributions to transform or "warp" the variables to standard normal, then impose a

multivariate normal structure based on the correlation matrix. This procedure is independent of the individual warping functions, which may be different for different individual variables, and in particular, can incorporate our market-based option distributions for individual variables representing securities with active options markets. A slightly different approach is to use correlations of the warped variables. This procedure is likely to be more accurate, but may involve more computational time.

We indicate some details. As before, it is notationally convenient to use v_i as a second notation for x_i, favoring the latter for fixed values and the former as a variable. Let C be the historical correlation matrix of the log variables $(\ln v_1, \ldots, \ln v_n)$, whose entries are the cross-correlations

$$\rho_{ij} = \frac{\mathsf{E}(\ln v_i \ln v_j) - \mathsf{E}(\ln v_i)\mathsf{E}(\ln v_j)}{\sqrt{\mathsf{E}(\ln v_i)^2 \mathsf{E}(\ln v_i)^2}}.$$

Then all diagonal terms ρ_{ii} are equal to 1, and C is a positive semi-definite covariance matrix, which we may here assume to be nonsingular (positive definite). If we use instead correlations of warped variables, we have simply

$$\rho_{ij} = \mathsf{E}(y_i y_j).$$

Let us define $F_C(y_1, \ldots, y_n)$ as the cdf of a multivariate Gaussian random variable with mean zero and covariance matrix C. Thus, $F_C(b_1, \ldots, b_n)$ is the probability that each variable y_i is at most some value b_i. There are more elaborate versions, such as $F_C(a_1, \ldots, a_n; b_1, \ldots, b_n)$, giving the probability that each y_i satisfies $a_i \leq y_i \leq b_i$. In the single-variable case these latter functions are obtained from the simple cdf by a single subtraction, involving two terms, but the corresponding bivariate case involves four terms, and in n dimensions there would be 2^n terms. However, each of these more elaborate cdf's can be directly computed as an integral, just like the simple cdf. Since the more elaborate cdf's are needed for Monte Carlo calculations, possibly in high dimensions, it is best to think of them as being computed directly.

We then define the multivariate cdf's

$$F(x_1, \ldots, x_n) = F_C(y_1(x_1), \ldots, y_n(x_n)),$$

and

$$F(a_1, \ldots, a_n; b_1, \ldots, b_n) = F_C(y_1(a_1), \ldots, y_n(a_n); y_1(b_1), \ldots, y_n(b_n))$$

where the $y_i(x_i)$ are the known warping functions for the individual variables. We find it convenient, with some abuse of language, to speak of $F(x_1, \ldots, x_n)$ as "the cdf", even though we have all of the above functions in mind, and to use $F(x_1, \ldots, x_n)$ as a proxy for the whole distribution (which it does, theoretically, determine). This multivariate cdf then has the following properties:

- Since the marginals of $F_C(z_1, \ldots, z_n)$ are Gaussian with mean 0 and variance 1, the marginals of $F(x_1, \ldots, x_n)$ are equal to $N(y_i(x_i)) = F(x_i)$; *i.e.*, they are correct according to each single-variable model.

- If the logvariables $(\ln v_1, \ldots, \ln v_n)$ are actually jointly Gaussian, then the multivariate cdf $F(x_1, \ldots, x_n)$ is correct.

In summary, the true joint distribution is approximated by a jointly lognormal distribution using historical correlations, combined with warping functions on each variable such that the marginal distribution of each variable is correct according to a selected single-variable model

(for example, according to our single-variable model for optionable securities, or according to the lognormal model using historical volatility). The single variables may actually be portfolios, with a default distribution for the portfolio return being lognormal, based on historical volatility. This multivariate theory generalizes both our single-variable theory and standard multivariate (log)Gaussian models. It again allows for market input through option prices, to the extent that components have an active option market, but does not exclude nonoptionable securities, and also allows portfolios as single variables. In this way BARRA (or functionally equivalent) factors are also allowed because of their interpretation as portfolios of long and short positions.

4 Applications to portfolios

Given the multivariate cdf $F(x_1, \ldots, x_n) = F_C(y_1(x_1), \ldots, y_n(x_n))$, we can answer many typical questions. We first give an overview, and then take up some of the applications in more detail.

As one example, suppose that we want to find the cdf of a portfolio variable

$$x = h_1 x_1 + h_2 x_2 + \cdots + h_n x_n,$$

where the h_i are arbitrary coefficients. A simple Monte Carlo method, probably not the fastest, is to draw random samples from the jointly Gaussian distribution with cdf $F_C(y_1, \ldots, y_n)$, transform each y_i via the inverse mapping function $x_i(y_i)$, and then compute the resulting output sample

$$x = h_1 x_1(y_1) + \ldots + h_n x_n(y_n).$$

After enough samples, we will have an approximation to the cdf of x. More precisely, the probability that $a \le x \le b$ is, approximately, the average number of samples $y_1, \ldots, y_n$ with $a \le h_1 x_1(y_1) + \ldots + h_n x_n(y_n) \le b$, and this approximation becomes exact in the limit for large sample sizes. This works for real portfolios, or for portfolios constructed from a number of assets and a residual variable, as might arise from a regression. Usually the regression is done in the log domain, which we discuss below. Note that the Monte Carlo method just described works perfectly well if the expression for x above is replaced by any function $f(x_1, \ldots, x_n)$ of the x_i, possibly quite nonlinear.

4.1 Log domain portfolios

In this subsection, we point out how our methods fit with another paradigm in common use in the financial community, and set up some further notation. It is common to work in the return domain, or equivalently, with logarithms; *i.e.*,

$$\ln x = \beta_1 \ln x_1 + \cdots \beta_n \ln x_n.$$

Ignoring any possible identification of these variables with those in the previous section, the same discussion and Monte Carlo method as above applies, if we regard x as a nonlinear portfolio $x = f(x_1, \ldots, x_n) = \exp(\beta_1 \ln x_1 + \cdots \beta_n \ln x_n)$. If the sum B of the β_i's is 1, such an x may be written $x = \widehat{h}_1 x_1 + \widehat{h}_2 x_2 + \cdots + \widehat{h}_n x_n$ where $\widehat{h}_i = \beta_i x / x_i$. Even if B is not 1, incremental changes ("returns") $d \ln x$ computed from this equation for x are consistent with the above expression for $\ln x$. It is common in the financial community to think of $\widehat{h}_i$ as approximately a constant h_i, so

that for short periods, where the $x_i's$ do not change too much, this equation for x is comparable to the portfolio equation in the previous subsection.[2]

For an asset x not given explicitly in terms of the terms of the x_i, we obtain a similar expression via linear regression:

$$\ln x = \beta_0 \ln x_0 + \beta_1 \ln x_1 + \cdots + \beta_n \ln x_n,$$

The β_i for $i \neq 0$ are correlation coefficients chosen to minimize the variance of the residual in historical data (perhaps subject to constraints, such as $\beta_i \geq 0$ and $\sum_{i=1}^{n} \beta_i = 1$). For example, x might be an security without a quoted option, and the x_i for $i \neq 0$ could be taken as assets for which we individually know the probability distributions, in addition to the required correlation coefficients for x. We have written the residual term as $\beta_0 \ln x_0$ (usually thinking of $\beta_0 = 1$ and the residual as normally distributed).[3] The mean of the latter could be nonzero, giving the regression "alpha"— a constant term making the mean of the regression correct. Alternatively, we could modify the equation to allow an explicit alpha, and keep the residual mean zero. Another minor variation might include the addition of a dummy variable with constant return, to adjust the value of x up or down. In particular, this gives another way of adjusting the residual mean to zero. This equation gives the previous one as a special case if we allow $\beta_0 = 0$.

4.1.1 Fast fits of portfolios

One approach, which promises to be relatively fast computationally, is the following. As in the development of cdf-implied volatilities in Section 2, let us assume that each logvariable $\ln x_i$ above is "Gaussian" with nonconstant variance $\sigma_1(x_i)^2 T$. In other words, the cdf is given by $F(x_i) = N(-d_2(x_i, \sigma_1(x_i), T))$. Our aim will be to give $F(x)$ by a similar equation, using some kind of fitted curve $\sigma_1(x)$. We will assume that we have some class of volatility curves in mind, with a small number of parameters which must be determined.

If the variables $\ln x_0, \ldots, \ln x_n$ were truly jointly Gaussian, then $\ln x$ would also be Gaussian. Its variance would be given by the formula

$$\mathrm{Var}(\ln x) = \sum_{i,j} \beta_i \sigma_i \rho_{ij} \beta_j \sigma_j T,$$

where ρ_{ij} is the correlation between $\ln x_i$ and $\ln x_j$, and $\sigma_i^2 T = \mathrm{Var}(\ln x_i)$. We therefore define the estimate $\hat{\sigma}_1(x)$ of $\sigma_1(x)$ by the conditional expectation

$$\hat{\sigma}_1(x) = \mathsf{E}(\beta_i \sigma_1(x_i) \rho_{ij} \beta_j \sigma_1(x_j) \mid \ln x = \sum_i \beta_i \ln x_i).$$

The calculation of the above conditional expectation may be done with Monte Carlo methods. In the language of nonlinear portfolios above, we would take the function $f(x_1, \ldots, x_n)$ to be 0 outside a thin multidimensional solid enclosing the hyperplane defined by $\ln v = \sum_i \beta_i \ln v_i$). Inside the solid we would take $f(x_1, \ldots, x_n)$ equal to the above expression for $\mathrm{Var}(\ln x)$, divided by the probability of being in the solid (also a Monte Carlo calculation). In terms of samples, we just take the average of $\mathrm{Var}(\ln x)$ over all the samples that end up inside the thin solid. However,

[2]Thus $\hat{h}_i \frac{dx_i}{x} = \beta_i \frac{dx_i}{x_i} = \beta_i d \ln x_i$, so that for small changes dx_i the change dx from the first equation is approximately the same as would be obtained from the second. However, this relationship requires "rebalancing" to remain a good approximation for longer periods.

[3]For the residual term $i = 0$, we can use a constant variance, or impose some generic nonconstant structure based on observed behavior.

it is not necessary to compute all values of $\hat{\sigma}_1(x)$, but only enough to fit the parameters for the volatility curves we are using.

The estimated mean of $\ln x$ would be $\ln s^* - \frac{1}{2}\hat{\sigma}_1(x)^2$, with s^* determined as before, or replaced with some risk-averse estimate, to obtain the risk-averse or "true" distribution. (It is common, incidentally, to use factor models such as these to estimate a risk-averse version of $\ln s^* = \sum_i \beta_i \ln s_i^*$ from risk-averse values of s_i^*.)

Also, we mention here one useful variation: We may prefer not to view the residual term $\beta_0 \ln x_0$ as part of the model, and instead write down a joint pdf only for $\ln x_1, \ldots, \ln x_n$. In this case we can use the double expectation

$$\hat{\sigma}_1(x) = \mathsf{E}(\mathsf{E}(\sum_{i,j} \beta_i \sigma_1(x_i) \rho_{ij} \beta_j \sigma_1(x_j) \mid \ln x = \sum_i \beta_i \ln x_i)),$$

where the inner expectation is with respect to the variables $x_1, x_2, \ldots, x_n$, and the outer expectation is with respect to the residual. We might take the standard deviation $\sigma(x_0)$ of the residual (taking $\beta_0 = 1$) as a constant, determined historically, or make an estimate based on some leverage model.

Now we can estimate the cdf $F(x)$ by

$$\hat{F}(x) = N(-d_2(x, \hat{\sigma}_1(x), T))$$

as in the univariate case. To summarize, we use our multivariate model to determine parameters for a univariate model of the portfolio. After that is done, we can obtain probabilities for the portfolio without having to go back to the multivariate model, thus achieving a savings in time. We could take this one step further and think of randomly generating values of $\sigma_1(x, T)$ independently of any Monte Carlo philosophy (but perhaps still throwing away values of x too far out-of-the-money), and then using the values obtained to do the regression required in the Dumas-Fleming-Whaley approach.

5 "What-if" questions

The multivariate distribution lends itself to the study of many questions regarding conditional probabilities. For example, suppose that we want to know the effect of the increase or decrease of some segment of the market on a portfolio, or the increase or decrease of some macro-economic factor. BARRA, following earlier ideas of Ross, has viewed such macro-economic factors as portfolios with both long and short positions. Similarly, BARRA considers market segments associated to price-to-earnings ratios and other fundamental parameters, as well as to industry groupings, as portfolios. (See the book of Grinold-Kahn cited above.) Thus, we are led simply to consider the effect of one portfolio on another.

For definiteness, let us suppose the first portfolio is x, where as above

$$\ln x = \beta_0 \ln x_0 + \beta_1 \ln x_1 + \cdots + \beta_n \ln x_n,$$

and the second portfolio is y, where

$$\ln y = \gamma_0 \ln y_0 + \gamma_1 \ln x_1 + \cdots + \gamma_n \ln x_n.$$

We take $\beta_0 = \gamma_0 = 1$, and view $\ln x_0$ and $\varepsilon = \ln y_0$ as residuals with mean 0. The latter residual is not assumed to be a factor in our multivariate model. Consider the following typical "what-if"

question: Let A and B be given positive constants. If we know $x \geq A$ at time T, what is the probability that $y \geq B$ at time T? We give two approaches to this problem, the first probably quicker, but possibly not as accurate, using a regression to avoid at least some Monte Carlo calculations.

5.1 "What-if": An approach involving part regression, part Monte Carlo

We have $\ln y \geq \ln B$ iff $\ln y - \varepsilon \geq \ln B - \varepsilon$. All correlations ρ_{ij} between $\ln x_i$ and $\ln x_j$ are assumed known. We may also assume that we have historical values of volatilities $\sigma_i = \sqrt{\mathrm{Var}(\ln x_i)}$. (Alternatively, we could estimate such values as expected values of implied volatilities, but it would not be difficult to maintain an inventory of historical values, and more in the spirit of this part of the calculation to do so.) Thus we can estimate the historical covariances between $\ln x$ and $\ln y - \varepsilon$:

$$\mathrm{Cov}(\ln x, \ln y - \varepsilon) \approx \sum_{i,j} \beta_i \sigma_i \rho_{ij} \gamma_j \sigma_j T,$$

as well as $\sigma_{\ln x} = \sqrt{\mathrm{Var}(\ln x)}, \sigma_{\ln y - \varepsilon} = \sqrt{\mathrm{Var}(\ln y - \varepsilon)}$ and the correlation

$$\rho(\varepsilon) = \rho_{\ln x, \ln y - \varepsilon} = \frac{\mathrm{Cov}(\ln x, \ln y - \varepsilon)}{\sigma_{\ln x} \sigma_{\ln y - \varepsilon}}.$$

This gives a standard regression for the variable $\ln y - \varepsilon$ expressed in standard deviations from its mean, in terms of a similarly standardized expression for $\ln x$. Note that ε has mean 0 by construction. Put $d_2(s^*, x, \sigma) = \frac{\ln(s^*/x) - \sigma^2 T/2}{\sigma\sqrt{T}}$. Thus $-d_2(s^*_x, x, \sigma_{\ln x})$ measures standardized $\ln x$ using historical volatility, and $-d_2(x) = -d_2(s^*_x, x, \hat{\sigma}_1(x))$ measures "standardized" (warped) $\ln x$ using the cdf-implied volatility curve $\hat{\sigma}_1(x)$, as discussed in the previous section. Here s^*_x denotes our best estimate for the value of x at time T.

Let $\sigma_1(y, \varepsilon)$ denote the volatility curve associated with $\ln y - \varepsilon$, which may be estimated as in the previous section (or computed from estimates of $\sigma_1(y)$ and the standard deviation of the residual, if we are willing to view the residual as uncorrelated with $\ln y - \varepsilon$, as is guaranteed in unconstrained regression). Put $d_2(y, \varepsilon) = d_2(s^*_y, ye^{-\varepsilon}, \hat{\sigma}_1(y, \varepsilon))$, so that $-d_2(y, \varepsilon)$ is a "standardized" measure of $\ln y - \varepsilon$. Then the standard regression appropriate to our model is

$$-d_2(y, \varepsilon) = \rho(\varepsilon)(-d_2(x)).$$

There is a residual associated with this regression, which we have not written down. It is presumably normal, and its variance may be computed. For notational reasons we will just imagine it has been incorporated into the original ε. As is apparent from the form of the expressions in the display, an alternative to the above regression is to do it with the warped correlation coefficients suggested in the previous section. If, in addition, it was appropriate to view the original portfolios as linear combinations of warped variables (our standard normal marginals), the regression above could be done without any recourse to Monte Carlo calculations. Similar remarks would apply if we used constant historical volatility functions throughout, though presumably the latter procedure would lose accuracy.

In any case, we can now answer our "what-if" question as a simple expectation in the univariate normal distribution of the (adjusted) residual ε. Abbreviate $d_2(s^*_x, A, \hat{\sigma}_1(A))$ to $d_2(A)$ and

$d_2(s_y^*, Be^{-\varepsilon}, \hat{\sigma}_1(B,\varepsilon))$ to $d_2(B,\varepsilon)$. Assume $\rho(\varepsilon) > 0$ (the natural case of a positive correlation). Then we have

$$\begin{aligned}
\Pr\{y \geq B \mid x \geq A\} &= \mathsf{E}(\Pr(-d_2(y,\varepsilon) \geq -d_2(B,\varepsilon) \mid -d_2(x) \geq d_2(A)) \\
&= \mathsf{E}(\Pr(-d_2(x) \geq \rho(\varepsilon)^{-1}(-d_2(B,\varepsilon)) \mid -d_2(x) \geq -d_2(A))) \\
&= \mathsf{E}(\min\{1, N(\rho(\varepsilon)^{-1}(d_2(B,\varepsilon))/N(d_2(A))\}).
\end{aligned}$$

The first equation follows just because $-d_2(y,\varepsilon)$ is monotonically increasing as a function of y; that is, the condition that $y \geq B$ is completely equivalent to the condition $-d_2(y,\varepsilon) \geq -d_2(B,\varepsilon)$. Similar remarks hold for the condition $x \geq A$, while the expression $\Pr\{y \geq B \mid x \geq A\}$ just means the probability that the condition $y \geq B$ holds when it is known that $x \geq A$. The second equation is then derived with the displayed expression above for $-d_2(y,\varepsilon)$. (If $\rho(\varepsilon)$ is negative, the inequality involving its inverse reverses.) This inner expectation is then calculated in the normal distribution. For values of ε for which $-d_2(A))$ is as large as $\rho(\varepsilon)^{-1}(-d_2(B,\varepsilon))$, the expectation is a certainty, and yields the value 1. When $-d_2(A))$ is smaller than $\rho(\varepsilon)^{-1}(-d_2(B,\varepsilon))$, its cumulative normal distribution value $N(-d_2(A))$ is smaller than $N(\rho(\varepsilon)^{-1}(-d_2(B,\varepsilon))$, and the probability $1 - N(-d_2(A)) = N(d_2(A))$ that the standard normal variable $z = -d_2(x)$ is at least $-d_2(A)$ is smaller than the corresponding probability $1 - N(\rho(\varepsilon)^{-1}(-d_2(B,\varepsilon)) = N(\rho(\varepsilon)^{-1}(d_2(B,\varepsilon))$ that z be at least $\rho(\varepsilon)^{-1}(-d_2(B,\varepsilon)$. The ratio $N(\rho(\varepsilon)^{-1}(d_2(B,\varepsilon))/N(d_2(A))$, which is the desired inner expectation, is thus smaller than 1, as is appropriate for a probability, conditional or not. If $\rho(\varepsilon)$ is negative, similar reasoning leads instead to the expression $\mathsf{E}(\max\{0, (N(d_2(A)) - N(\rho(\varepsilon)^{-1}(d_2(B,\varepsilon)))/N(d_2(A))\})$ for the desired conditional probability. Although the final answer in either case is an expectation (over ε), it is essentially an integral that could be computed quickly with power series. (A very simple and accurate power-series expansion of $N(z)$ is given on p. 252 of the book by Hull cited above.) Using that, one could determine by iterative methods what value of ε makes, say, the ratio $N(\rho(\varepsilon)^{-1}(d_2(B,\varepsilon))/N(d_2(A))$ equal to 1, and then integrate the ratio against the standard normal pdf from $-\infty$ to the determined value of ε,in the $\rho(\varepsilon) > 0$ case. Similar remarks apply if $\rho(\varepsilon) < 0$. (Note that, if $\rho(\varepsilon) = 0$, the variables $\ln x$ and $\ln y$ are uncorrelated, and the conditional probability $\Pr\{y \geq B \mid x \geq A\}$ is the same as the unconditional probability $\Pr\{y \geq B\}$.)

All of the latter calculations can be done very fast. Of course, we have already used some Monte Carlo calculations to get this far, unless we are in the simplified context of constant volatility functions.

5.2 "What-if": The full Monte

It is easy to say how we would compute an answer to the same "what-if" question, using our full joint probability distribution. We simply write

$$\Pr\{y \geq B | x \geq A\} = \mathsf{E}(\Pr(-d_2(y,\varepsilon) \geq -d_2(B,\varepsilon) \mid -d_2(x) \geq -d_2(A))$$

and interpret $\ln x$ in $-d_2(x)$, and $\ln y - \varepsilon$ in $-d_2(y,\varepsilon)$ in terms of their expansions in $\ln x_0$, $\ln x_1, \ldots, \ln x_n$. To compute, say the inner expectation by a Monte Carlo calculation, we would generate a large number of random samples of multivariate standard normal vectors z with covariance matrix C. We then take the average, over the samples z which happen to satisfy $z \geq -d_2(A)$, of the function which is 1 when $-d_2(y,\varepsilon) \geq -d_2(B,\varepsilon)$ and 0 otherwise. We have not experimented to see whether this method yields better answers than the regression procedure above. Nevertheless, it illustrates how we could approach more sophisticated "what-if" questions

that could not be easily treated by regressions. For example, suppose we believe that factor w will remain in a range $C \leq w \leq D$, and ask the same question about y, subject to the same condition on x. This is hard to formulate in terms of regression, and is simply not possible in terms of single-factor regression. However, it is easy to answer with the full distribution:

$$\Pr\{y \geq B \mid x \geq A, C \leq w \leq D\} =$$
$$\mathsf{E}(\Pr(-d_2(y,\varepsilon) \geq -d_2(B,\varepsilon) \mid -d_2(x) \geq d_2(A)), -d_2(C) \leq -d_2(w) \leq -d_2(D)).$$

Finally, we may not want to work in the log domain, which, if we started with a fixed portfolio $x = h_1x_1 + h_2x_2 + \cdots + h_nx_n$ would force us into an approximation, as noted. But, working with the full distribution, we can phrase a condition $x \geq A$ as $h_1x_1(y_1) + \ldots + h_nx_n(y_n) \geq A$, in the language of the first section where the vector of $y's$ plays the role of our vector z here. Monte Carlo calculations can now proceed as before, using log domain expressions or not for the other conditions.

6 "You gotta believe" questions

In the previous section we were focusing on an investor thinking about the value of his or her portfolio y in response to the change in a factor x. Conversely, an investor might want to know what the investment world looks like if a given stock or index y goes to a certain level B at time T. What is the expected value A at time T of another portfolio x, or simply of one of the factors x_i? Our main plan is, upon input by the user that y is going to level B, to list several assets x_i or factors/indices x most highly correlated with y and their expected values with y at B.

It would also be possible to display a confidence interval for each selected asset or factor, and have other information about its new projected probability distribution readily available. We could also offer comparisons with the old projected probability distribution of x, where no assumptions on y is made. Finally, in some cases, where it was possible to explain much of the variance of y with just a few x_i (appearing in the regression of y), we could list percentage increases/decreases of a portfolio of these x_i required to make B the expected value of y, based solely on its dependence on this portfolio. (For example, the coefficients in the portfolio could come from the regression of y with respect to all the x_i, or some new regression might be done, perhaps allowing user-defined constraints). It should be mentioned that medians or modes are alternatives to expected values (means) here and above; in any case users will need to be educated about the fact that the median and mode differ systematically from the mean in near lognormal distributions.

The main problem might be viewed as understanding the probability distribution of x, given that $y \geq B$ at a given time T, with x and y as in the previous section. This can be approached by the methods of the previous sections, by reversing the roles of the variables.

There is, however, a simpler question that can be treated in an especially quick way. Consider the problem of determining the mean of x conditioned on the equality $y = B$ at time T. The idea is to use simple regression methods, but interpret answers as measured in terms of our variable volatilities. In our previous notation, we have a regression

$$-d_2(x) = \rho \cdot (-d_2(y,\varepsilon)) + v$$

where ρ (which we called $\rho(\varepsilon)$ in the previous section) is the historically determined correlation between $\ln x$ and the random variable $\ln y - \varepsilon$. Note that the roles of dependent and independent

variable are reversed. There is also a residual v, which has mean 0 here, and plays no role (gets averaged away). Thus, the desired conditional expected value A of x is obtained from

$$-d_2(A) = E(-d_2(x) \mid y = B) = \rho \cdot E(-d_2(s_y^*, Be^{-\varepsilon}, \widehat{\sigma}_1(y, \varepsilon))) = \rho \cdot (-d_2(s_y^*, B, \widehat{\sigma}_1(y, \varepsilon)).$$

Recall that $\widehat{\sigma}_1(y, \varepsilon)$ is an estimate, obtained by Monte Carlo methods, of the implied volatility σ_1 associated to the random variable $\ln y - \varepsilon$. For faster but less accurate calculations it can be estimated historically as $\sum_{i,j} \beta_i \sigma_i \rho_{ij} \beta_j \sigma_j$ with each of the σ's, β's, and ρ's here given historically. (See the previous section for notation.) Similarly, for fast calculations, $-d_2(x)$ could use historical volatility, though we expect it to be given more accurately, or rather, more accurately according to the market view, as $-d_2(x) = -d_2(s_x^*, x, \widehat{\sigma}_1(x))$, using the implied volatility function estimate $\widehat{\sigma}_1(x)$. If $x = x_i$ is a single asset or index in our model, then $\widehat{\sigma}_1(x) = \widehat{\sigma}_1(x_i)$ does not require a Monte Carlo estimate, but is presumably already available.

To summarize, the conditional expected values required to answer "you gotta believe" questions are easily obtained by regression methods. The accuracy of such answers is enhanced, or at least shaped more to reflect market input, when all logvariables are measured in "standard deviations," interpreted as our variable volatilities.

7 Portfolios containing option securities

We conclude this document by briefly pointing out that our methods, when using full Monte Carlo calculations, easily apply to portfolios containing option securities. The well-known idea is to think of an option as as a kind of nonlinear portfolio— a quadratic one, to be more precise. Thus, an option on a single underlying security with underlying price x_1 has a price approximately $x = c + \Delta(x_1 - s_1) + (1/2)\Gamma(x_1 - s_1)^2$ for x_1 near s_1, where the option was evaluated to a known value c. Here Δ and Γ are well-known parameters in the options markets, giving the first and second derivatives of the option price at s_1 with respect to the underlying security price x_1. Perhaps the most characteristic feature of options is that they have nonzero Γ— their proportion of increase or decrease with respect to the underlying security price changes as the security price changes. Explicit formulas in terms of other standard parameters are available, say, in the Black-Scholes theory for both Δ and Γ (see the Hull book cited above). Such formulas could be obtained by differentiation directly in other theories or when using empirically-fitted curves. In any case, once we have such an explicit approximation to x, its probability distribution is easily given by the Monte Carlo methods of Subsection 3.1 above. The same method applies as well to portfolios containing several options and other securities.